DOCUMENT RESUME

ED 380 063 IR 016 984

AUTHOR Daniels, Lee
TITLE Audio-Vision: Audio-Visual Interaction in Desktop Multimedia.
PUB DATE [95]
NOTE 8p.; In: Imagery and Visual Literacy: Selected Readings from the Annual Conference of the International Visual Literacy Association (26th, Tempe, Arizona, October 12-16, 1994); see IR 016 977.
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Articulation (Speech); *Audiovisual Aids; Authoring Aids (Programming); Cognitive Processes; *Computer Mediated Communication; Computer Software; *Imagery; *Multimedia Materials; Music; *Noise (Sound); Sound Effects; *Theory Practice Relationship

ABSTRACT

Although sophisticated multimedia authoring applications are now available to amateur programmers, the use of audio in of these programs has been inadequate. Due to the lack of research in the use of audio in instruction, there are few resources to assist the multimedia producer in using sound effectively and efficiently. This paper addresses the problem by providing some basic understanding of the cognitive and affective effects of audio when used with visual material in a computer-mediated environment, and presents some general tips on choosing and manipulating audio elements. Topics include: foundations of multimedia; general properties of sound and perception; and audio elements in multimedia production: speech, sound effects, and music; silence; the roles of audio: picture defines sound, sound defines picture, sound parallels picture, and sound counterpoints picture. Contains 14 references. (DGM)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

# Audio Vision: Audio-Visual Interaction in Desktop Multimedia

Lee Daniels

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Alice Walker
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED 380 063

Until recently, the resources needed to author multimedia computer programs required an abundance of time, advanced programming skills, and expensive hardware, relegating the development of multimedia products to professional developers and highly skilled researchers. Recent advancements in computer technology have produced desktop hardware and software that is more powerful, flexible, affordable, and user-friendly than ever before. Sophisticated multimedia authoring programs that are easy to use (i.e. Hypercard™, Digital Chisel™, etc.) allow the non-programmer to author multimedia productions that rival the best of commercial products. These new authoring applications provide text, visual, and sound resources which can be incorporated into a rich computer mediated environment by both educators and students. As a result, many resources are available to guide the amateur programmer in designing multimedia programs and improving visual presentations.

Audio, on the other hand, has been almost an afterthought. Due to the lack of research in the use of audio in instruction there is are few resources to assist the multimedia producer in using sound effectively and efficiently in multimedia (Thompson, Simmonson, & Hargrove, 1992). As a result, most authors either utilize "stock" sounds that are "thrown in" without contemplating or understanding the relationship between audio and visuals, or ignore the audio medium altogether. Too often, a sound effect or music segment is used solely as a device to gain attention, and not as an integral part of the multimedia message.

This paper attempts to address this problem by providing some basic understanding of the cognitive and affective effects of audio when used with visual material in a computer mediated environment. In addition, some general tips on choosing and manipulating audio elements are presented.

## The Foundations of Multimedia

Although the term multimedia has been used for years by both educators and industry, there is little agreement on an exact definition (Strommen & Ravelle, 1990). For the purposes of this paper the term multimedia refers to a computer mediated environment that incorporates two or more media types such as images (still or moving), text, graphics, sound, and other data.

The effectiveness of multimedia as an instructional medium is based on the theory of multiple-channel communication. Multiple-channel communication involves synchronous presentation of information "...through different sensory channels (i.e., sight, sound, touch, etc.) which will provide additional stimuli reinforcement" (Dwyer, 1978, p. 22). The benefits of adding additional media channels when communicating is offered by Severin's (1967) cue summation theory which asserts that learning will be increased when stimuli that share information are presented because they reinforce each other. An alternate view of communication assumes that there is only one channel of communication, and that additional cues across channels offer no advantage, and run the risk of "overloading" the human processing system (Travers, 1964). In an attempt to reconcile the two theories, Hsia (1968) hypothesized that communication through multiple channels could be more effective so long as the central nervous system was not overloaded. In their

IR016984

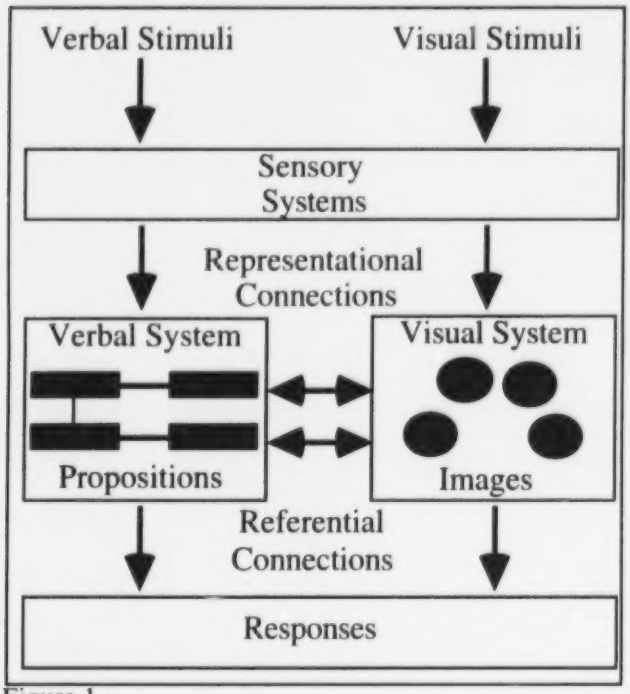

Figure 1

review of multimedia literature, Moore, Myers, & Burton, (in press) suggest that when *highly related* cues are summated across channels, multiple-channel presentations are superior to single channel presentations .

Another integral component of multiple-channel communication is Allan Paivio's dual coding theory (Paivio, 1991). Paivio's theory is based upon the assumption that memory and cognition are served by two separate symbolic systems, one specialized for dealing with verbal information and the other with images. Pavio (1971) defines an image as "...nonverbal memory representations of concrete objects and events, or nonverbal modes of thought" (p. 12). Paivio distinguishes images from verbals which relate to speech or a language system (Paivio, 1971). While visual stimuli is normally associated with images, other modalities (i.e. auditory) may also produce images. Although each system can function separately, most processing involves connections between the two systems (see figure 1). The word "car" for instance, may translate into images of cars, and likewise, a visual of a car may form the verbal symbol (word)"car". Paivio points out that although words *can* be imaged, images are associated with verbals automatically, which would explain the superiority effect of visuals (Pressley & Miller, 1987).

As noted earlier, both text and speech are received as verbals. In addition, concrete sounds can be interpreted as images. Thus, dual coding is not effective when the information sources are coded within the same mechanism (Barron & Atkins, in press) This concept is important to remember when one uses audio in conjunction with visuals. Narration with text *does not* constitute an additional channel, and *can* cause intra-channel interference. Likewise, combining sound effects and visuals presents the same danger.

To summarize, it appears that when information is presented across channels, it should be highly correlated to improve learning and avoid inter-channel interference. Additionally, The multimedia author should ensure that images or verbals presented across channels are not conflicting.

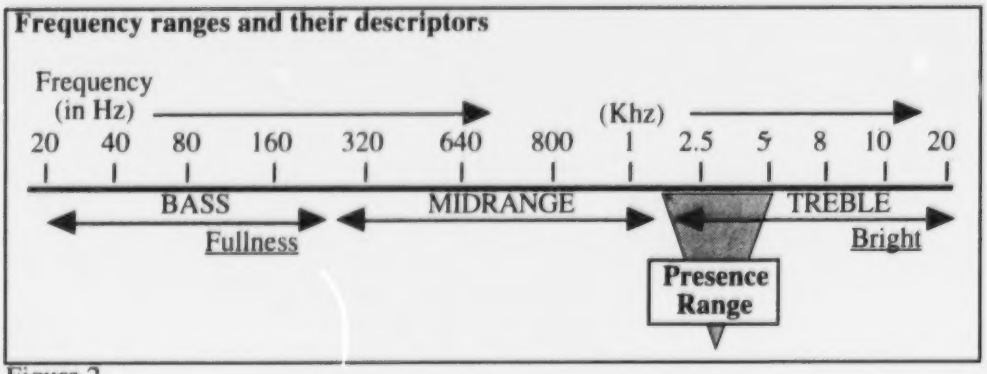

Figure 2

## General Properties of Sound and Perception

Sound is perceived by the brain as a description of the object that emits the sound *and* the environment that the sound occurs in. The sonic *quality* of a sound allows the listener to make judgments about the spatial location, relative size, and environment of the sound (Runstein & Huber, 1986). When combined with visuals, these characteristics can enhance or confuse the visual message depending on how the aural and visual attributes relate to one another. All sound elements share the characteristics that define location. The spatial location of a sound is perceived by tone quality, relative volume, and amount of reflections (echoes/reverb).

Tone quality refers to the brightness or dullness of the sound. Sounds that are farther away have less high frequency information (treble) than those that are closer. Sounds farther away contain more low frequencies (bass). Our hearing apparatus is most sensitive to sounds in what is known as the presence range (see fig. 2). By amplifying a sound's spectral content in this 2kHz-5kHz range, the audio appears to be closer, louder, and more intelligible. Additionally, sounds that are farther away tend to have a lower pitch than those that are closer (Rossing, 1982).

Sounds that are closer, obviously sound louder than those at a distance. When placing a sound/image source at a distance it should be audible, but to some degree softer than those placed in the foreground. As noted above, sound sources that are at a distance tend to have more reflections or echoes, than those nearby (Rossing, 1982). In addition to defining space, these same characteristics also identify location and movement. In order to define movement, stereo sound is desirable, wherein the intensity of the sound changes as its source moves laterally. Even in mono however, a change in volume, tone quality, and pitch can enhance the perception of movement. When using sound effects in multimedia, many authors overlook the importance of matching the aural and visual characteristics of space and location. A wolf visually placed in the foreground conflicts with the sound of a far away howl. Similarly, moving objects whose sound is static can cause confusion in the viewer. This is especially important when cognitive information is being supplied by the audio track. Often in animations depicting physical, mechanical, or scientific properties, the sound of the action or process can serve as a valuable cue to the information being presented. The sound of a malfunctioning part is often as important as the image itself.

The size of a sound source is usually intimated by pitch and echoes. A deep pitched voice with lots of echoes would imply a large, perhaps ominous being, while high pitched canine sounds conjure up images of puppies rather than large dogs. When a sound is placed inside an environment, the reflections should match the environment. A cavernous pit would entail many echoes, while a classroom interior would have little if any echoes.

Sound also contributes to the pace of visual presentation. Narration, dialogue,

sound effects, or music can establish a fast and hectic pace or a slow and somber mood to compliment the visual. Although often overused, unique audio elements are excellent ways to gain and maintain attention throughout a multimedia program. Musical interludes, odd sound effects or different voices can alert the viewer to pertinent visual elements or serve as a reminder to stay on task. These same devices can also be used to signify transitions to new topics or themes. Recurring music and sound effects are often used to identify characters, events, or places as when a rattle identifies the villain, or a happy melody signifies the protagonist (Alten, 1990).

**Audio Elements in MultiMedia Production**

There are only three audio elements that the multimedia developer has to work with: 1) Speech, 2) sound effects, and 3) music. Initially, this may seem to comprise a small arsenal of communicative tools when compared to the multitude of visual elements (color, texture, angle, etc.) available to the multimedia designer, especially considering the superiority of visual memory. However, one only has to look at the effectiveness of radio as a medium for communicating cognitive and affective information, or to the excellent use of audio in many documentaries such as Ken Burn's "The Civil War" to realize that sound has a substantial impact in a "visual" medium.

**Speech**

Speech can function as either narration or dialogue. Narration, like text, is often used to deliver concrete information. When presented with text however, narration should be highly redundant since both text and speech are perceived as a verbal proposition. Any dissonance between the two channels can distract the user, cause interference, and result in less retention, or misinterpretation. Therefore, narration is most useful as a replacement *of* text and not as addition *to* text.

One instance when narration is more appropriate than text is when screen space is at a premium, and additional text would reduce the visual impact (i.e. a complex scientific display or detailed artwork) (Barron & Atkins, in press). Narration is also preferred when it is necessary to direct the viewers attention to details of the visual. Consider a program dealing with artistic details and attributes. Text often is used to guide the viewer to specific attributes (figure 3). Narration would allow the viewer to concentrate on the visual image rather than moving from text to visual. In addition, the focal point of the screen, the image, is able to occupy more screen space (figure 4).

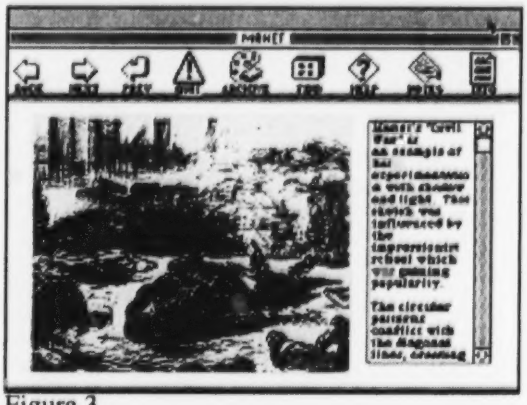

Figure 3

Figure 4

The pace of narration or dialogue can also heighten the intensity of the visual. A fast moving narration adds to the intensity of time lapsed animation, for instance, while a slow, steady narration compliments the somber



BEST COPY AVAILABLE

mood of a hero's funeral. Similarly, fast paced dialogue between two characters can reflect tension, anger, excitement, or nervousness. Smooth, even paced dialogue reflects friendliness, relaxation, and confidence.

As noted earlier, the tone quality of the narration can have an effect on the listener's perception. A narration that is bright and present is perceived to be closer, and therefore, more intimate and trustworthy as opposed to speech that sounds dull and distant (Alten, 1990). Also, a narration track with an amplified presence range will be more intelligible and require less volume to be heard over music or sound effects (Woram, 1989).

**Sound Effects**

We usually think of sound effects as being contextual, literally interpreting the visual as it appears. Such is the case of a dog's bark, the roar of a jet engine and the like. However, they can also serve a narrative function by adding more to a visual's apparent information (Alten, 1990). Descriptive effects contribute to the subtle, sonic aspects of an image. For instance the sound of gentle ocean surf may include gulls, people playing, and boat sounds to set a particular mood Commentative sound also tells more about an image, usually unrelated to the visual itself. Imagine a program about air pollution, and a scene of city traffic. Treating and blending the car engines to sputter and "cough" comments on the detriment to the air we breathe.

**Music**

Perhaps no other sound is as effective as music in communicating complex emotions and moods. Music, can define a locale with ethnic melodies. It can establish time with musical elements that suggest a period in history such as the 1960's or the Roman era, etc. Music can identify characters and events with recurring themes. as well as providing transitions from one idea to another. Varying tempo and rhythm, contributes to pace can provide counterpoint to the visual to create tension and irony.

**Silence**

Probably the most underrated sound element is silence. Silence can have enormous affective impact, especially when it is unexpected. Silence creates tension simply by letting the user's imagination "fill in" the sound. The aftermath of a plane crash or the disappearance of a character in a story both lead to suspense as to what will happen next. Another excellent use of silence to build tension (and a staple of horror movies) is to slowly remove sound elements one at a time so that the absence of sound "sneaks up" on the viewer.

**The Roles of Audio**

When audio and visuals are presented concurrently, the audio-visual relationship takes on dynamics and meanings that are different than when either media is presented alone. When combined with visuals, audio assumes one of four roles: 1) Picture defines sound, 2) sound defines picture, 3) sound parallels picture, and 4) sound counterpoints picture.

**Sound Defines Picture**

Imagine a multimedia program on the Brazilian rain forest. A still image of the jungle interior is accompanied by the solitary sounds of the environment: rainfall, bird calls, and other animals, along with lively ethnic music. Alternatively, the sound element could contain the sound of chain saws, machinery, and ominous music. Two different interpretations of the rain forest are implied.

**Picture Defines Sound**

The sound is defined when the visual image is so strong, that the accompanying sound is a literal translation of the image. A raging tropical storm, with crashing waves and bent palms nearly demands a soundtrack that consists of wind, surf, and rain sound effects. Audio is supportive of the dominant visual, reinforcing the image.

**Sound Parallels Picture**

This is the most common relationship between audio and visual elements. In this relationship the audio element combines with the visual element to create a mood or deliver information that is more potent than either element alone. The sounds of battle with gunshots, cannon, and anguished screams compliments the visual of a battle scene. The ferocity and destruction of war is conveyed by both media separately, but is intensified by both elements together.

**Sound Counterpoints Picture**

When sound counterpoints picture, both media contain unrelated information that creates an effect that is not conveyed by either media alone. For example, in a presentation on the civil rights movement, irony is created when a visual montage of segregated public facilities is underscored by a reading of the United States Constitution.

**Summary**

Many affordable sound editing programs (such as Macromedia's SoundEdit™) give the multimedia author the power to manipulate sound files in nearly any way imaginable. By adjusting tone, pitch, volume, duration, and other sound characteristics, sound files can be fine tuned to fit the specific visual element. Using these techniques, the same sound file can treated in many ways and used as several sound effects. Narration or dialogue can be edited to establish pace. Voices can be altered via special effects to establish identities, and allow the same voice talent to become several characters. Increasing or decreasing musical tempo and pitch is also possible.

The examples mentioned above are intended only as guides only - not hard and fast rules. The key to finding the best audio for a particular visual is to experiment and listen carefully. As with any other aspect of multimedia design, the audio element is somewhat intuitive and changes with each new situation. Beta versions should be developed and assessed before deciding on the final product. It is also good practice to pay close attention to the multimedia programs that you admire, or ones that are successful, chances are they have high quality, well designed audio elements. Try to discern what the audio elements are, how they were produced, and how they relate to the visual. Keep a log of devices and techniques that work and adapt them for your programs - there's no need to reinvent the wheel!

## References


Alten, S. R. (1990). Audio in media. Belmont, CA: Wadsworth

Barron, A. E., & Atkins, D. (in press). Audio instruction in multi-media education: Is textual redundancy important? Journal of Educational Multimedia and Hypermedia.

Chester, G., Garrison, G. R., & Willis, E. E. (1971). Television and radio. Englewood Cliffs, NJ: Prentice-Hall.

Dwyer, F. M. (1978). Strategies for improving visual learning. State College, PA: Learning Services.

Moore, D. M., Myers, R. J., & Burton, J. K. (in press). Multiple-channel communication: The theoretical and research foundations of multimedia. In Jonnassen, D. (Ed.), Handbook of research on communication and educational technology. Washington, D.C.: Association for Educational Communications and Technology.

Paivio, A. (1991). Dual coding theory: retrospect and current status. Canadian Journal of Psychology, 45 (3), 255-287.

Paivio, A. (1971). Imagery and verbal processes. New York: Holt, Rinehart & Winston.

Pressley, M. & Miller, G. (1987). Effects of illustrations on children's listening comprehension and oral prose memory. In D.M. Willows & H.A. Houghton (Eds.), The psychology of illustration. New York: Springer-Verlag.

Rossing, T. D. (1982). The science of sound. Reading, MA: Addison-Wesley.

Runstien, R. E., & Huber, D. M. (1986). Modern recording techniques. Indianapolis, IA: Macmillan.

Severin, W. (1967). Another look at cue summation. AV Communication Review, 15(3), 233-245.

Strommen, E. F., & Revelle, G. L. (1990). Research in interactive technologies at the Children's Television Workshop. Educational Technology Research and Development, 38(4), 65-80.

Thompson, A. D., Simonson, M. R., & Hargrave, C. P. (1992). Educational technology: A review of the research. Washington, D.C.: Association for Educational Communications and Technology.

Woram; J. M. (1989). Sound recording handbook. Indianapolis, IA: Macmillan.